INTERACTIVE REAL-TIME CONCATENATIVE SYNTHESIS IN VIRTUAL REALITY

Carl Moore and William Brent

American University
Audio Technology Program
4400 Massachusetts Ave NW
Washington DC, USA
carlmoore256@gmail.com, w@williambrent.com

ABSTRACT

This paper presents a new platform for interactive concatenative synthesis designed for virtual reality and proposes further applications for immersive audio tools and instruments. TimbreSpace VR is an extension of William Brent's TimbreSpace software using the timbreID library for Pure Data. Design and implementation of the application are discussed, as well as its live performance aspects. Finally, future work is laid out for the project, proposing versatile audio manipulation software specifically for XR platforms.

1. INTRODUCTION

Concatenative synthesis techniques typically use a large database of sounds that are segmented into smaller units or grains for synthesis of new sounds. These grains are analyzed to obtain descriptors or attributes pertaining to their timbre, which allows the grains to be reorganized into a new sound known as a "target." In "Free synthesis," the user manually selects audio grains for real-time playback rather than using an automated system [1]. In short, concatenative synthesis is a platform for creating dynamic soundscapes, unconventional musical performances, and novel sound effects. These ends have been achieved through software implementations such as CataRT, and previous iterations of TimbreSpace by William Brent [2]. This paper introduces a new platform for concatenative granular synthesis and audio analysis, implemented in virtual reality.

Virtual and augmented reality (XR) mediums provide exciting new platforms to experience sonic information in the visual domain that has previously been confined to the two dimensions of a screen. Although virtual reality is rapidly becoming a popular platform for unique types of creation and interaction, few interactive synthesis tools have emerged for platforms such as the Oculus Rift or HTC Vive. Mux, a modular synthesizer available Steam VR, is a notable example; however, as of this paper, it still remains in an early access stage of development.

TimbreSpace VR is a tool that provides a palette of timbres encoded as discrete sonic events that can be easily located, patched together, and rearranged in new ways. The primary goal is to create a versatile sonic workspace that provides visual and haptic feedback for sonic exploration and creation.

2. DESIGN

TimbreSpace is a synthesizer which relies largely on bark-frequency cepstral coefficients (BFCCs) for distribution of audio grains in 3D space. BFCCs are a subset of cepstral analysis, a general process of reducing the complexity of spectral analysis results. BFCCs perform well in timbral analysis because they are based on a frequency scale that closely coincides with human frequency perception relative to critical bands [3].

2.1. Interface

The application presents a cloud of grains derived from a given pre-analyzed audio source and provides the user with two wands as the primary means of interaction. Grains are positioned in the 3D space according to any three of the descriptors derived from analysis in PureData using timbreID. Users can specify the X, Y, and Z spatial ordering of grains upon initialization, as well as set the world scale and grain sphere size multiplier. This flexibility is necessary for different use case scenarios.

Grains appear as spheres which are scaled individually according to amplitude. By default, grains are spaced within the scene according to their 1st, 2nd, and 3rd BFCCs, which are meaningful representations of timbre similarity. Each grain is colored according to their 4rd, 5th, and 6th BFCCs, providing yet another dimension of timbral visualization. Therefore, grains of similar color and location will sound similar.

The aesthetic decisions for the interface are largely carried over from William Brent's original implementation of TimbreSpace. The location and color of grains are simple, easy to understand indicators of timbre, and several first-time users expressed a strong understanding of these tonal-visual correlations. Additional features of grain scaling which were not present in the original version of TimbreSpace also help to further indicate the status of a grain before hearing it play.

The overall aim of the visual aesthetics is to communicate tonal characteristics of regions within the grain cloud so that the user can focus on conceptualizing a tone in their mind and find it quickly, rather than searching for the sound they want by audition.

2.2. Controls

In previous real time concatenative synthesis applications, user interaction was very much constrained within the spatial dimensions. William Brent's open-air fingertip navigation was an improvement to existing one-dimensional gestures available in concatenative synthesizer applications such as CataRT, utilizing IR

https://doi.org/10.21785/icad2019.068

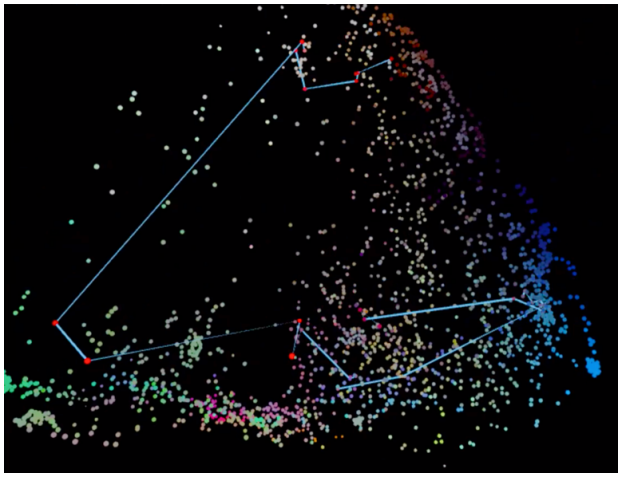

Figure 1: A grain cloud comprised of around 2000 electronic drum samples

fingertip tracking as a means of interaction with audio grains [4]. In addition, gestural combinations such as pinch and rotation of the hands enabled the modulation of effects parameters such as delay and transposition [4].

TimbreSpace VR extends the concept of wrist rotation as an effects modulation parameter, however, gestural tracking is now gathered from the Oculus Touch controller's built in gyroscope and accelerometer.

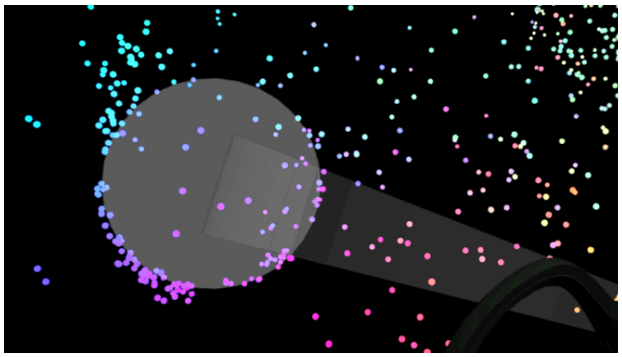

Figure 2: Grains colliding against the bubble cursor wand

Two separate "wands" with spherical tips are provided for playing and sequencing the grains. They use the various features of the Oculus Touch controllers to accomplish all the actions available within TimbreSpace VR. The main goal is to allow the user to reach any grain visible within the scene through the use of natural, intuitive controls, while maintaining precise sonic control.

The spherical wand tip is a three-dimensional implementation of a bubble cursor, a GUI selection tool which possesses benefits from accuracy by dynamically resizing [5]. The diameter can be resized via the combination of a control button and a "doorknob twist" style action, allowing for quick and accurate moves which are useful for live performances. Both wand tips can also be repositioned along the Z-axis independently, allowing the user to reach closer or further into the grain cloud.

2.3. Workflow

Audio scenes or "soundscapes" are currently loaded into the Unity editor as a collection of text files (containing attribute data,) and mono .wav files, which can then be read by TimbreSpace VR. To generate these assets, preliminary steps involving normalization and silence removal, and then audio analysis for feature extraction in PureData are required to import new soundscapes. The process is not particularly user-friendly at this stage but can easily be integrated into a single package with further work.

Soundscapes are typically generated with an artistic concept in mind, or out of curiosity. Collections of thousands of flute samples, bird calls, or rain storms are just a few examples of content that has been experimented with. These samples are often organized first in a handful of ways. Dynamic scenes can be created by combining sonic elements that evoke certain imagery into a DAW session. The result is a nonlinear audio scene in which a performer, much like a Foley artist, acts out the sonic scene within TimbreSpace. Because the process of timbre analysis removes the temporal component entirely, clips can be sewn together in random orders. TimbreSpace also works well with collections of instrument samples and loops, which can be navigated and played as an instrument.

Samples of a percussive nature are often put through a custom PureData patch before analysis, in order to optimize feature extraction in timbreID. This stage utilizes the Bark~ object within timbreID to detect transients and concatenate thousands of samples into a specified window for further analysis.

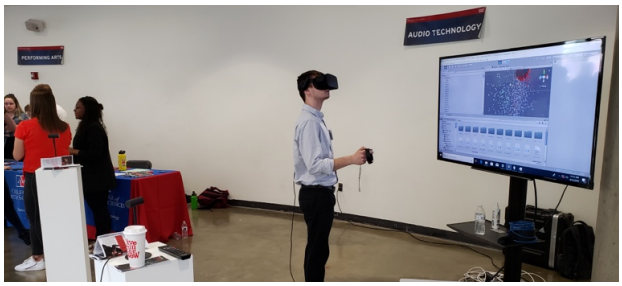

Figure 3: A user exploring a grain cloud in TimbreSpace VR

Figure 4: A large constellation seen from a distance in the scene

2.4. Constellations

Sequencing of audio events is one the primary new innovations of TimbreSpace VR. Grain sequences (referred to as Constellations) are dynamic groups of audio segments that are looped and played back in a specified order. They are displayed as lines running from one grain in the loop to the next, graphically indicating the trajectory of the sound through the scene. Using the constellation editor, the user can form musical ideas in live and non-real-time scenarios. The ordering of grains within the 3D space based on audio features allows for sounds to blend together in novel, timbrally coherent arrangements.

Constellations allow for the creation of musically discrete patterns and rhythms, making TimbreSpace VR a space for dynamic musical composition. They take a completely different interactive and visual approach to musical sequencing when compared to traditional musical sequencers, leading to abstract and unexpected phrases, loops, and textures. The spatial workflow shows the potential of XR applications to bring new tools to artists and contribute to new forms of creative expression.

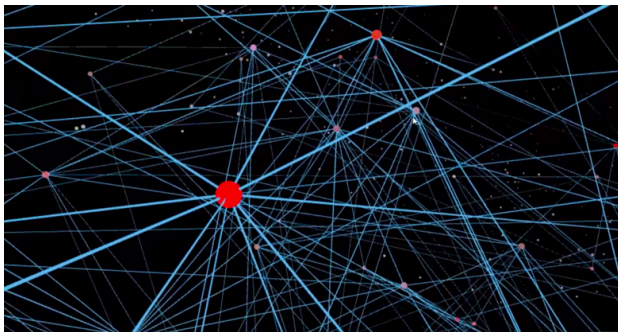

Figure 5: A constellation up close

2.5. Physics

Simulated physical response of audio grains provides a basis for many interactions unique to TimbreSpace. The kinetic response of grains attempts to introduce dynamics to a performance that provide feedback to the performer and audience in the form of action-sound relationship cues. Each unit is pinned to a given position in 3D space with an elastic bond. Interactions via the two bubble wands can stretch these links upon intersection with a grain, displacing all grains within the wand's diameter. Grains bounce about the surface of the wand causing new collisions with the wand as well as with other grains, activating those sonic events in random orders. The result is a controlled method of exciting grain clusters, creating unpredictable concatenation patterns that avoid audibly looping.

3. TECHNOLOGY

The core application is built using the Unity game engine and written in C#. TimbreSpace is being developed for the Oculus Rift headset and controllers. Audio descriptor metadata is currently generated externally using William Brent's timbreID library in PureData, and exported as a text file containing each of the 26 attributes (descriptors) for each grain event.

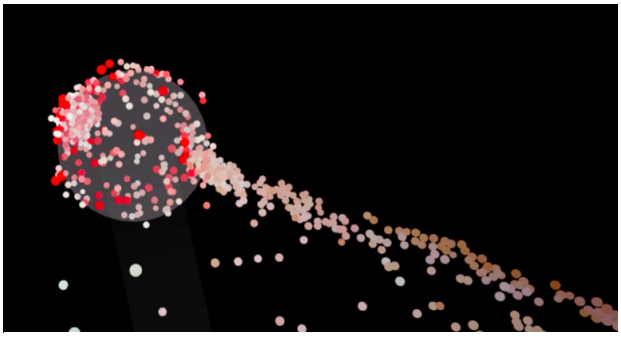

Figure 6: The bubble cursor's simulated physical interaction with audio grains

In this implementation of timbreID, the patch segments a given audio file and measures various audio attributes within each frame. These attributes include frame number, pitch, amplitude, harmonicity, spectral centroid, and BFCCs. Window size has a significant impact on the sonic results of concatenative synthesis, therefore timbreID employs several different window sizes for its analyses, ranging from 2048 to 16384 samples in length. This provides flexibility in TimbreSpace VR, which has the ability to load a scene at a variety of different grain sizes. Because the exported database contains a list of attributes for each grain, any of this information can be referenced in the scene.

Running on current graphics hardware (Nvidia GTX 1070), TimbreSpace VR can render scenes with grain clouds up to around 5000 grains without significant performance reduction or framerate stutter. This equates to around 15 minutes of audio content if played back at a grain size of 8192 samples, an ideal length for musical exploration.

4. FUTURE WORK

The full realization of TimbreSpace is a fully modular sonic sandbox for artists. A collection of tools will extend the capabilities of the control set so that any audio event or cluster of audio events can be added, removed, modified, networked and sequenced. These events will include audio grains, filters, and logical operators which will be exposed to a patching network. Alongside this, greater functionality will be added to the live performance controls to allow precise real-time synthesis.

Additional GUI controls currently being implemented will greatly expand the functionality of TimbreSpace VR. Constellations will be able to be duplicated, saved, and recalled, as well as played polyphonically alongside any number of other active sequences. Information about each grain (determined during analysis) will be accessible in the interface, making TimbreSpace VR a useful platform for visual audio analysis. The ultimate goal is to provide a space for sonic exploration and experimentation that feels natural and provides new tools for artistic expression.

5. REFERENCES

[1] Schwarz, Diemo. "Concatenative Sound Synthesis: The Early Years" Journal of New Music Research35, no. 1 (2006): 3–22. doi:10.1080/09298210600696857.

[2] Diemo Schwarz, Gregory Beller, Bruno Verbrugghe, Sam Britton. “Real-Time Corpus- Based Concatenative Synthesis with Catart”, *Expanded version 1.1 of submission to the 9 Int. Conference on Digital Audio Effects*

[3] Brent, William. “Perceptually Based Pitch Scales in Cepstral Techniques for Percussive Timbre Identification”.

[4] Brent, William. “Physical Navigation of Virtual Timbre Spaces with TimbreID and DILib”, Proceedings of the 18th International Conference on Auditory Display, 2012.

[5] Grossman, Tovi, and Ravin Balakrishnan. “The Bubble Cursor.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems — CHI 05, 2005. doi:10.1145/1054972.1055012.